(Sociocentric) Network Survey Data Collection

13-Sep-2006

Agenda

- **Nuts and bolts**
  - Boundary specification
  - Sampling
  - What to measure
  - Survey media and formats

- **Research issues**
  - Gaining access
  - Informant accuracy
  - Key informant approaches
  - Ethnographic sandwich
  - Ethical issues

Bounding the Survey

- Extremely vexing to beginners and outsiders
  - Network concept would seem to argue against boundaries
    - Empirical research makes clear we are all connected
  - Even if distant links don't matter, some people in the sample will be at the edge, no matter where we cut it
- One key is to isolate when bounding matters
  - Yes: Interpersonal influence studies
  - No: selection studies
- In snowballs, need to remove edge

Types of Boundaries

- Realist (emic) vs nominalist (etic)
- Attribute-based
  - Top management team at Enron
  - Drug injectors in Hartford
- Relation-based
  - Snowballing out from seed sample until few or no new names (i.e., exhaust current component)
  - But is a component a real boundary?
- Mixed criteria
  - Sexual ties among residents of Nang Rong
- Theoretical criteria

Sampling

- Sampling is not a problem for ego networks

- Sampling for complete networks is in its infancy
  - Snowballing especially problematic
  - Certain measures ok, such as density

- New work in adaptive sampling may help

Ego Networks

- (Random) sample of nodes
  - Each sampled node called an "ego"
- Each is asked for set of contacts called "alters"
- Ego also asked (usually) about ties among alters
- Connections between ego's or between alters of different egos are not recorded
  - Each ego is a world in itself
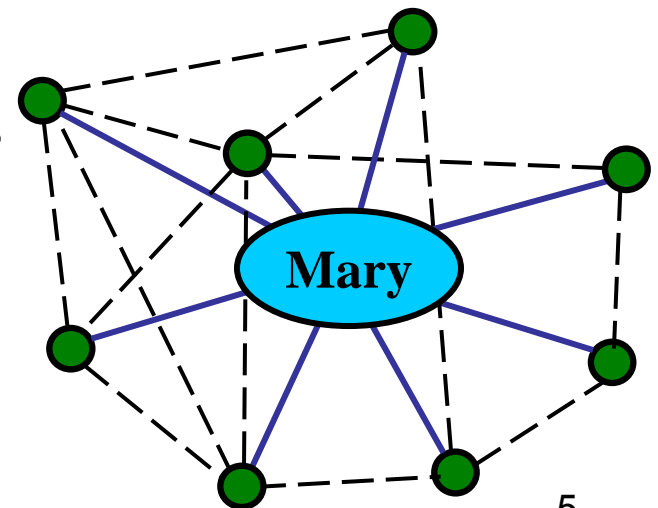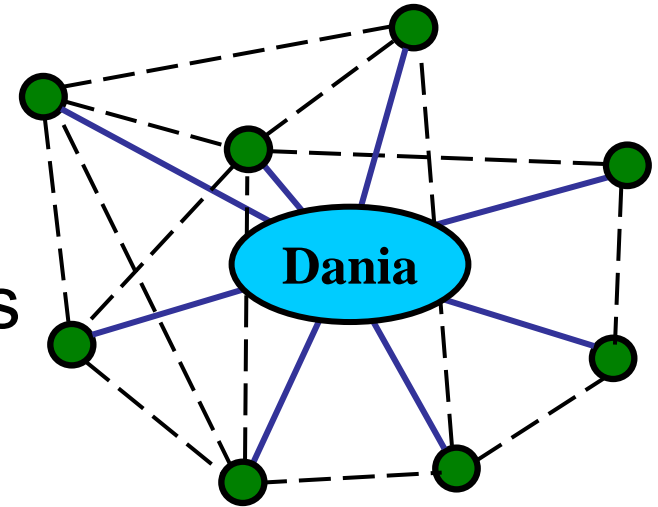
Ego Network Surveys

- ## Name generator
  - Obtain complete list of alters
    - Exact names not needed

- ## Name interpreter
  - Systematic assessment of social relations with each alter

- ## Alter attributes
- ## Alter-Alter ties

(back to sociocentric networks)

Which relations to measure?

- Inquiring minds want to know: Which questions best tap **the** group's network?
  - Looking for validated "social network scale"
- But "The" network does not exist.
  - People connected by a multitude of social ties of different types.
  - Researcher gets to measure whichever relation(s) are appropriate for the research goals
  - Each type of tie (e.g., social relation) forms a network with its own structure and meaning
- A social relation is just a variable. "giving advice" is to network analysis what "attitude toward gun-control" is to survey research.
  - In attitude research, do you ask which attitudes are best to survey?
- It is the researcher who defines the relations of interest
  - But is true that measuring emically non-salient relations can be challenging
  - E.g., Check off the people who send Christmas cards to your friends
  - Who do you communicate with more by phone than by email?

Survey Choices

Confidentiality Reminder

- This is in addition to consent form

Social Network Questionnaire

Thanks for participating. Please note that the data generated in this survey are NOT anonymous and are NOT confidential. The results will be used in the workshop in Washington. **Important note: you must enter your name in Question 0.**

When you're done, press the "Submit" button. Thanks for your help.

Q0. What is **your name**: [   ]

Closed-Ended vs Open-Ended

Roster of names or just blank lines?

- ## Closed-ended (aided)
  - Requires bounded list
  - Can be impractical for large networks

- ## Open-ended (unaided)
  - Subject to recall errors
  - Can limit number of choices made (more effort, limited space)

Name	Q1. Heard of them
Allata, Joan	☐
Baer, Justin	☐
Baker, Ted	☐
Bercuwitz, Rick	☐
Branzei, Oana	☐
Brooks, Scott	☐
Brower, Ralph	☐

If you wanted to get something done on behalf of a customer who would you contact? *(write as many names as you like in the spaces provided)*

_____ _____

_____ _____

_____ _____

_____ _____

Hybrid Questionnaire

1. If you wanted to get something improved or done on behalf of a customer who would you contact?

    Name   (index no.)

    _Denny Terio_____ (169)

    _Eric Estrada_____ (27)

    _____ (   )

    _____ (   )

2. If you wanted to get a true reading on where [company name] was headed as an organization, who would you talk to?

    _____ (   )

    _____ (   )

> Paper version uses separate booklet containing name directory

> Web version uses drop-down menus

Repeated Roster vs MultiGrid

Q1. Please indicate which of the following you had met or been aware of before coming to this workshop.

Allata, Joan	☐
Baer, Justin	☐
Baker, Ted	☐
....	☐

Q2. Check of f the names of the people you know. By "know" I mean that you have spoken to each …

Allata, Joan	☐
Baer, Justin	☐
Baker, Ted	☐
....	☐

Q1. Using the checkboxes below, please indicate **who you have heard of or know about** among the participants of the workshop.

Q2. Check off the names of the **people you know**. By "know" I mean that you can attach a name to a face, you have spoken to each other at least once, and the other person is also likely to put you down.

Q3. Check off the names of people you **have worked with** on a paper or other academic/administrative project.

Q4. Check off the the names of a selected set of people whom you don't know but **would like to know**, based on things you've heard, or their interests, etc.

Name	Q1. Heard of them	Q2. Know them	Q3. Worked with	Q4. Want to know
Allata, Joan	☐	☐	☐	☐
Baer, Justin	☐	☐	☐	☐
Baker, Ted	☐	☐	☐	☐
Bercuwitz, Rick	☐	☐	☐	☐
Branzei, Oana	☐	☐	☐	☐
Brooks, Scott	☐	☐	☐	☐
Brower, Ralph	☐	☐	☐	☐

Question Wording Issues

- "Friendship" does not mean the same thing to everyone
  - Especially across national cultures
- Some helpful practices:
  - Use one word label plus two or three sentence description, plus have full paragraph detailed explanation available
  - Don't make find distinctions
    - Liking, friendship, esteem, respect, feel positive towards
  - Use homogeneous samples

Tick or Rate?

- Ask resp for yes/no decisions or quantitative assessment?
  - Yes/no are cognitively easier on resp (therefore reliable, believable),
  - Yes/no *much* faster to administer
  - But yes/no provides no discrimination among levels
- A series of binaries can replace one quant rating:
  - Instead of "How often do you see each person?"
    - 1 = once a year; 2 = once a month; 3 = once a week; etc.
  - Use three questions (in this order):
    - Who do you see at least once a year?
    - Who do you see at least once a month?
    - Who do you see at least once a week?

Paper or Plastic?

- Paper medium
  - Reliable
  - Reassuring to respondents
  - Errors in data entry
  - Data entry is time-consuming
- Electronic
  - Span distances, time zones
  - Harder to lose
  - Fewer data handling errors
  - Lower response rate
  - Emailed documents vs survey instruments

Dillman Design Considerations

- Network questionnaires can be fun but are usually time-consuming and generate anxiety

- Providing value

- Treating resp with respect

- Attractive formatting

- Cloaked in authority and importance

Coping with common data problems

Unexpected Asymmetry

- M claims to have sex with B, but B does not claim to have sex with M
    - The relation is logically symmetric, but empirically asymmetric
    - errors of recall; strategic response
- Sometimes asymmetry is the point
- Logically symmetric data may be symmetrized
    - if either A or B mentions the other, it's a tie
    - Only if each mentions the other is it a tie

Non-Symmetric Relations

- Gives advice to
- Can't symmetrize logically non-symmetric relations, except by changing meaning of tie
- Unless you ask question both ways:
  - Who do you give advice to?
  - Who gives advice to you?
- Two estimates of the A$\rightarrow$B tie, and two estimates of the A$\leftarrow$B tie

Missing Data

- For logically symmetric relations
  - if Xij is missing, substitute Xji
  - If whole row missing, substitute corresponding column

- For logically non-symmetric relations, ask questions both ways (who do you give advice to, who gives advice to you)
  - set Aij = Bji
  - i.e., missing row is replaced with column of the inverse relation

Survey Construction Strategies

- Row-based (for undirected relations)
- Row and Column-based (for directed relations)
- Matrix based (Krackhardt CSS)

22

Row-Based

- Each informant questionnaire corresponds to one row in the network adjacency matrix

- Issues of comparability across respondents

- For logically undirected relations, can deal with accidental asymmetry and missing respondents via symmetrization
  - Intersection rule: $X_{ij} = 1$ if $X_{ij} = 1$ and $X_{ji} = 1$
  - Union rule: $X_{ij} = 1$ if $X_{ij} = 1$ or $X_{ji} = 1$

Row and Column Based

- Each informant effectively asked to fill out both their row and their column of the adjacency matrix (but actually stored as separate matrices)
  - Aij: Who do you give advice to?
  - Bij: Who do you get advice from?
- Handle asymmetry by creating new matrix $X = A \cap B^T$ (intersection criterion)
  - $Xij = 1$ iff (Aij = 1) AND (Bji = 1)
  - i.e., i gives advice to j if i says i gives advice to j and j says they receive advice from i
- Problem with cognitive & affective relations – resp is expert

Matrix-based

- Krackardt CSS

- Each respondent asked about relations among all pairs of persons in group, not just those involving self
  - Yields network matrix $C(k)$ for each respondent

- Aggregate respondent matrices using choice of rules
  - Local: $X_{ij} = 1$ if $C(i)_{ij}$ and $C(j)_{ij}$
  - Global: $X_{ij} = 1$ if $C(k)_{ij} = 1$ for most k

Krackhardt CSS

Q1. How well the members of each pair know each other:									
	Response scale: Blank = They have never met. 1 = They are merely								
Knowledge	Aaron	Ali	Dan	Dave	David	Ed	George	Greg	Howard
Aaron									
Ali									
Dab									
Dave									
David									
Ed									
George									
Greg									
Howard									

Response bias

- Some respondents positively biased
  - Give big numbers in general when rating strength of tie or frequency
- Row-based approach yields matrices in which each row potentially has different measurement scale
  - Can create asymmetry when none "exists"
- For valued data can normalize by rows
  - Z-scores, euclidean norms, maximum, marginals

Informant Accuracy

- Bernard, Killworth et al compared observed with recalled interaction data
  - Ham radios, deaf TTYs
  - About half of the cells in the adjacency matrix were wrong
- Romney & Faust noted that structural analyses didn't seem so far off
  - Surface structure vs deep structure
- Freeman, Romney & Freeman
  - Resps biased toward long term patterns

Krackhardt CSS

- Many sources of inaccuracy
  - Recall and exaggeration of ties with high status people
  - Idiosyncratic understanding of the question
- Take "average" of everyone's perception of given dyad's relationship
  - Capitalize on social cognition (see Dawes)
  - Great for deliberately hidden relationships

Gaining Access

- A little harder than for ordinary studies
  - Strong preference for complete data
  - Respondent fears
  - Length of interview

- Quid pro quo helps but muddies the ethical waters

Ethnographic Sandwich

- Ethnography at front end helps to …
  - Select the right questions to ask
  - Word the questions appropriately
  - Create enough trust to get the questions answered

- Ethnography at the back end helps to …
  - Interpret the results
  - Can sometimes use resps as collaborators

Ethical & Strategic Issues

- What makes the network case especially challenging ethically?
- What are the dangers & to whom?
  - In academic setting
  - In management setting
  - In mixed situations
  - In national security setting
- What can we do about it?

Ethical Issues

- Respondents cannot be anonymous
- Missing data are troublesome
  - Creating incentive to downplay dangers
  - Results may be wrong (cf use of polygraphs by courts)
- Non-participants still included
  - And participants are like informers
- Outputs ideally show individual level data
- Pushes boundary of the professional
- Deceptively powerful
  - is still unknown; looks like research
- Quid pro quo arrangements with research sites
  - Management is hiring/firing based on "research" results

3-Way Disclosure Contract

- For research done in organizations
- Signed by management, the researchers, and each participant
- Clearly identifies what will be done with the data

Management Disclosure Contract

Study Authorization

This document authorizes Steve Borgatti and Jose Luis Molina to conduct a social network study at Management Decision Systems (hereafter "the company") during the period January 1, 2005 to March 1, 2005.

Rights of the Researchers

The data – properly anonymized so that neither individual nor the company are identified -- will form the basis of scholarly publications.

Rights of the Company

In addition, the researchers will furnish the company with a copy of all the data. The company agrees that these data will not be shared among the employees and will only be seen by top management. The company agrees that the data will not form the basis for evaluation of individual employees, but will be used in a developmental way to improve the functioning of the company.

Rights of the Participants

The participants of the survey – the people whose networks are being measured – shall have the right to see their own data to confirm correctness. They may also request a general report from the researchers that does not violate confidentiality of the other participants regarding what was learned in the study.

Truly Informed Consent Form
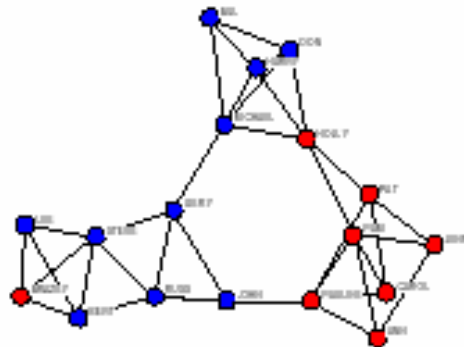

Page 1

35

Truly Informed Consent Form

Risks & Costs

Since management will see the results of this study, there is a chance that someone in management could consider your set of communication contacts to be inappropriate for someone in your position, and could think less of you. Please note, however, that the researchers have obtained a signed agreement from management stipulating that the data will be used for improving communication in the company and will not be used in an evaluative way.

Individual Benefits

We will provide you with direct, individualized feedback regarding your location in the social network of the organization.

Withdrawal from the Study

You may choose to stop your participation in this study at any time. If so, you will not appear on any of the social network maps and no metrics will be calculated that involve you. Note that management has agreed that participation in the study is voluntary.

Confidentiality

As explained above, your participation will not be anonymous. In addition, all of top management will be able to see results of the study that include your name. Outside of top management, however, the data will be kept confidential. Any publicly available analyses of these data will not identify any individual by name, nor identify the organization.

Participant's Certification

I have read and I believe I understand this Informed Consent document. I believe I understand the purpose of the research project and what I will be asked to do. I understand that I may stop my participation in this research study at anytime and that I can refuse to answer any question(s). I understand that management and only management will see the results of this research with individuals identified by name.
I hereby give my informed and free consent to be a participant in this study.

Signatures:

Page 2

36

The Dialectics of Data Collection

Using SNA for personnel decisions

→ Employees answer survey dishonestly

→ Researcher uses advanced data collection & analysis techniques

→ Employees collude

→ Employer relies on passive data such as e-mail

→ Employees communicate to look good; resentment

In the end, academic research suffers